

Comparing Contents of a Knowledge Base to Traditional Information Sources

Nunzia Bettinsoli Giuse¹, Dario A. Giuse^{1,2}, Richard A. Bankowitz¹, Randolph A. Miller¹.

1. Section of Medical Informatics, Department of Medicine, University of Pittsburgh

2. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA

Physicians rely on the medical literature as a major source of medical knowledge and data. The medical literature, however, is continually evolving and represents different sources at different levels of coverage and detail. The recent development of computerized medical knowledge bases has added a new form of information that can potentially be used to address the practicing physician's information needs. To understand how the information from various sources differs, we compared the description of a disease found in the QMR knowledge base to those found in two general internal medicine textbooks and two specialized nephrology textbooks. The study shows both differences in coverage and differences in the level of detail. Textbooks contain information about pathophysiology and therapy that is not present in the diagnostic knowledge base. The knowledge base contains a more detailed description of the associated findings, more quantitative information, and a greater number of references to peer-reviewed medical articles. The study demonstrates that computerized knowledge bases, if properly constructed, may be able to provide clinicians with a useful new source of medical knowledge that is complementary to existing sources.

INTRODUCTION

While the amount of information published in the medical literature expands, physicians' information needs often go unmet [1, 2, 3, 4, 5]. The medical literature forms the basis for much of the information used in daily practice. The medical literature, however, is not as complete and homogeneous an entity as commonly supposed. Any given medical concept is covered by numerous sources, not all equally accessible, each using slightly different nomenclature. Some sources report conflicting information, and most describe their information using only qualitative terms, whose interpretation can vary widely [6]. The task of the practitioner, investigator, or author is to sift through all of this information and distill from it pertinent, reliable knowledge.

An effective solution to physicians' growing information needs will require reliable, "distilled" information sources that summarize entire areas of medicine. Such a solution should allow physicians to access information at the level that is appropriate for each specific question. The system would resemble a pyramid, with the vast amount of primary medical knowledge lying at the base, and each higher level forming a synthesis of the levels below [7]. In addition to the actual medical data, the system would provide supporting evidence for each piece of data, since physicians are unlikely to accept statements not supported by objective clinical evidence.

Can the medical literature be synthesized into this type of structure, and to what extent will the results be rater dependent? To gain insight into these issues, we designed a series of studies aimed at clarifying the factors involved in acquiring and using medical knowledge. In the first of such studies [8] we documented how well a group of trained physicians agreed on the task of abstracting information from a fixed body of literature. In that study, seven academically based physicians independently created a comprehensive, detailed description for acute perinephric abscess. In doing so, they synthesized information from the literature into a formal structure known as a disease profile [9,10]. That study shed light on the process of literature interpretation by clinically active physicians, and provided a systematic evaluation of agreement among physicians who consulted the medical literature to extract precise diagnostic information. The study provided a first important result in the validation of the knowledge acquisition process, because it showed considerable agreement among physicians with diverse backgrounds at geographically distinct sites. The knowledge acquisition task we studied was complex, and took an average of two weeks full-time equivalent to complete.

The disease profile for perinephric abscess created during the first study is a condensed form of the material contained in 109 peer-reviewed journal articles. The participating physicians synthesized the

109 articles into 180 findings and 26 connections (or "links") to related diseases. Such findings and links, complete with quantitative information, describe the presentation of perinephric abscess in the QMR knowledge base.

The aim of the current study is to contribute to a better understanding of the relationship among different knowledge sources. In particular, the study is aimed at answering the following question: What information would practicing clinicians have readily available regarding perinephric abscess? In other words, how does the completed disease profile compare to existing information sources with a similar level of detail?

To answer this question, the study compares the information about perinephric abscess found in the disease profile with the information found in medical textbooks. While the methodology used to construct disease profiles is similar to meta-analysis [8], meta-analytic studies are not pervasive enough to provide coverage comparable to disease profiles or textbooks. Among the existing sources of medical information, therefore, textbooks are the most similar to disease profiles. Like disease profiles, textbooks present a synthesis of different opinions, medical articles, and other textbooks. Also like disease profiles, textbooks typically present aggregate information, rather than data about individual patients. The level of coverage, however, may differ substantially. To determine similarities and differences, we compared the information contained in the perinephric abscess disease profile with that contained in two general internal medicine textbooks and two specialized nephrology textbooks.

BACKGROUND

A QMR disease profile [9] consists of a list of findings and a list of links, i.e., connections to related diseases. Each entry is associated with numerical quantities known as the evoking strength (analogous to positive predictive value) and the frequency (analogous to specificity). The QMR knowledge base is a collection of more than 620 disease profiles which is used by the QMR medical decision support system.

A disease profile specifies in a precise and constrained vocabulary the history, physical, and laboratory findings reported in the literature for a given disease. In addition to this information, the

disease profile also contains pointers to the supporting evidence. Each entry is cross-referenced to the articles that were used to determine numerical data and to justify the inclusion of each finding or link.

Not all areas of medicine are covered as extensively in medical knowledge bases as in medical textbooks. The QMR knowledge base, for example, is primarily oriented towards supporting diagnosis, and as such it does not contain explicit pathophysiological information or treatment information. More generally, knowledge bases tend to concentrate on those areas that are most important for the underlying decision support systems. To make comparisons possible, therefore, in the remainder of this paper we will concentrate exclusively on those areas that are covered both by the QMR knowledge base and by typical medical textbooks.

METHODS

For this study, we selected both internal medicine textbooks and specialized textbooks. For the first category, we chose two widely used general internal medicine textbooks [11,12] which describe perinephric abscess at a level suitable for most internists. For the second category, we asked a faculty nephrologist to name representative nephrology textbooks that she might consult to review the findings associated with perinephric abscess. The nephrologist suggested two textbooks [13, 14].

The version of the disease profile used in the study was taken from the current QMR knowledge base. This version was the result of the final group review process [15], which is used any time a new disease profile is added to the knowledge base.

Each item mentioned in each textbook was checked against the disease profile. This process yielded the lists of findings and related diseases that appeared only in the disease profile, only in the textbook, and in both the disease profile and the textbook.

It quickly became apparent that one of the specialized nephrology textbooks [14] contained substantially more findings than the other three textbook sources. A more detailed comparison was performed between this textbook and the disease profile. Each finding in this textbook was classified according to whether it

was mentioned in the textbook without any quantitative information, mentioned with semi-quantitative information, mentioned with quantitative information, or mentioned in the textbook but not in the disease profile. When quantitative information was present, the information was compared against the one in the disease profile, and the agreement or disagreement was recorded.

RESULTS

Coverage of the entries reported in the disease profile varied greatly among the four textbooks, as shown in Table 1. The material on perinephric abscess in the two general internal medicine textbooks matched 20 and 52 entries (respectively) from the disease profile, corresponding to 9.7% and 25.2% of the entries in the disease profile. All entries in the two general textbooks were mentioned in the disease profile. The first specialized nephrology textbook matched 56 entries, which again were all mentioned in the disease profile. The second nephrology textbook, on the other hand, listed 3 entries that did not appear in the disease profile, and covered 128 (62.1%) of the 206 entries listed in the disease profile. This textbook listed a total of 131 entries.

Table 1: Number of entries in each textbook versus number of entries in the disease profile.

	General Textbook 1	General Textbook 2	Nephrol. Textbook 1	Nephrol. Textbook 2
Only in textbook	0	0	0	3
Only in profile	186	154	150	78
In both	20	52	56	128
textbook /profile	9.7%	25.2%	27.2%	62.1%

Among the three entries listed in the second nephrology textbook but not in the disease profile, one entry (Gallium Scan) had been omitted from the disease profile on purpose, because the profilers felt that the data presented in the medical literature were inconclusive. The other two entries were not reported

by any of the participants who had contributed to the disease profile.

Even the most comprehensive nephrology textbook contained considerably less quantitative information than the disease profile. Out of a total of 131 entries in the textbook, 100 did not contain precise information about the frequency of occurrence of the entry. Of these, 50 entries contained no quantitative information, and 50 contained only semi-quantitative information. Of the 31 entries that did contain quantitative information, 29 matched the frequency reported in the disease profile.

DISCUSSION

A variety of sources are available to address current physicians' information needs [16]. Traditional sources include other colleagues (still a very popular option), professional meetings, medical textbooks, and medical journals. More recently, other sources have become available which provide higher-level syntheses of information. Such sources include the ACP Journal Club [17], the Internal Medicine Alert [18], Journal Watch [19], the Year Book of Medicine [20], and others; consensus guidelines [21] such as those published by the NIH, AHCPR, AMA, ACP, and others; and, increasingly, computerized medical knowledge bases.

Despite the availability of these information sources, however, a central question remains to be answered: How well do we know what we really know? How does a clinician know that he or she can trust any or all of the existing sources, and what sources to use in the first place? Our analysis of the description of perinephric abscess in two general internal medicine textbooks and two specialized textbooks shows that some of the available information sources may, in fact, provide less complete coverage than generally expected. Not only did the textbooks report significantly fewer findings than the computerized disease profile; the number of findings in the textbooks having associated precise numerical information was also much smaller. While the textbooks covered areas of medicine that are not covered in a typical QMR disease profile, such as therapy, the disease profile described a much larger number of findings and provided more extensive quantitative information.

The development of computerized medical knowledge bases in recent years [22] has added a new source of

information to the repertory of the practicing physician. While the primary purpose of knowledge bases is to support computerized decision-making tools, it is evident that they could also be used to address the information needs of today's physicians.

As the present study shows, for certain topics knowledge bases may provide an effective complement to other sources of medical knowledge. In particular, they can be used as a complement to textbooks, which provide a similar level of information synthesis but different coverage of various aspects of medicine. The study, for example, shows that while the QMR knowledge base does not directly cover pathophysiology or therapy information, it provides substantially more complete coverage of the findings associated with a disease than even specialized nephrology textbooks. It would be very interesting to extend the study to other disease profiles within the knowledge base, and more generally to other knowledge bases, to determine systematically the relative coverage of different electronic knowledge bases versus different textbooks.

In addition to their complementary coverage, knowledge bases typically provide more quantitative information than textbooks. This emphasis on quantitative information is dictated by the use of knowledge bases for expert system decision support, which requires precise numerical data; textbooks are traditionally more oriented towards qualitative descriptions of disease states. Finally, electronic knowledge bases can provide very rich sets of pointers to the original evidence used in their construction. All recently created QMR disease profiles, for example, contain detailed bibliographical references for each entry. Such references allow the physician to bridge the gap between synthesized information (such as that contained in a knowledge base or in a textbook) and the primary medical literature that was used during the synthesis process. The nature of electronic storage makes it easy to provide extremely detailed references; this would be equivalent to having each line of a medical textbook contain five to ten bibliographical references.

Because of this potential for complementing existing information sources, and because of the virtually instantaneous access capabilities of the computer, the development of electronic medical knowledge bases represents a first step in the creation of a more integrated, more effective system of information sources. Clearly, much work remains to be done in

this area, ranging from the understanding of physicians' information needs to issues of verification and validation of large-scale computerized knowledge bases. The ultimate goal, however, is well worth the effort. If an effective, integrated "information pyramid" can be constructed, tomorrow's physicians will be able to access the medical information they need more efficiently and more effectively.

Acknowledgments

The research described in this document was supported by the National Library of Medicine, Research Grant no. R01 LM04622.

References

1. Jelovsek FR, Rittwage J, Pearse WH, Visscher HC. Information management needs of the obstetrician-gynecologist: A survey. *Obstet Gynecol.* 1989; 73(3):395-399.
2. Osheroff JA, Forsythe DE, Buchanan BG, Bankowitz RA, Blumenfeld BH, Miller RA. Physicians' Information Needs: Analysis of Questions Posed during Clinical Teaching. *Ann Int Med.* 1991; 114(7):576-581.
3. Forsythe DE, Buchanan BG, Osheroff JA, Miller RA. Expanding the Concept of Medical Information: An Observational Study of Physicians' Information Needs. *Comput Biomed Res.* 1992; 25:181-200.
4. Covell DG, Uman GC, Manning PR. Information Needs in Office Practice: Are They Being Met? *Ann Int Med.* 1985; 103:596-599.
5. Timpka T, Arborelius E. The GP's Dilemmas: A Study of Knowledge Need and Use During Health Care Consultations. *Meth Inform Med.* 1990; 29:23-29.
6. Kong A, Barnett GO, Mosteller F, Youtz C. How Medical Professionals Evaluate Expressions of Probability. *N Engl J Med.* 1986; 315(12):740-744.
7. Bernstein, L.; Siegel, E.; Goldstein, C. The Hepatitis Knowledge Base: a Prototype Information Transfer System. *Annals of Internal Medicine,* 1980, 93:169.

8. Giuse NB, Giuse DA, Miller RA, Bankowitz RA, Janosky JE, Davidoff F, Hillner BE, Hripcsak G, Lincoln MJ, Middleton B, Peden JG Evaluating Consensus Among Physicians in Medical Knowledge Base Construction. *Methods of Information in Medicine*, April 1993.
9. Miller RA, Pople HE, Myers JD. INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med*. 1982; 307:468-476.
10. Masarie FE, Miller RA, Myers JD. INTERNIST-I Properties: Representing Common Sense and Good Medical Practice in a Computerized Medical Knowledge Base. *Comput Biomed Res*. 1985; 18:458-479.
11. Wyngaarden JB, Smith LH. *Cecil Textbook of Medicine* (18th edition). W.B. Saunders Company, 1988.
12. Wilson JD, Braunwald E, Isselbacher KJ, et al. *Harrison's Principles of Internal Medicine* (12th edition). McGraw-Hill, Inc. 1991.
13. Brenner BM, Rector FC. *The Kidney*. W.B. Saunders Company. 1986.
14. Schrier RW, Gottschalk CW. *Diseases of the Kidney* (4th edition). Little, Brown and Company, Boston/Toronto. 1988.
15. Giuse NB, Bankowitz RA, Giuse DA, Parker RC, Miller RA. Medical Knowledge Base Acquisition: The Role of the Expert Review Process in Disease Profile Construction. In: *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care*. New York: IEEE Computer Society Press; November 1989.
16. Wyatt J. Use and sources of medical knowledge. *Lancet*. 1991; 338:1368-1373.
17. ACP Journal Club. American College of Physicians, Philadelphia, PA.
18. Internal Medicine Alert. American Health Consultants, Inc.
19. Journal Watch. The Massachusetts Medical Society. Waltham, MA.
20. Year Book of Medicine. Year Book Medical Publishers, Inc. Chicago - London - Boca Raton.
21. Audet AM, Greenfield S, Field M. Medical practice guidelines: current activities and future directions. *Ann Intern Med*. 1990; 113(9):709-714.
22. Miller RA, Giuse NB. Medical Knowledge Bases. *Acad Med*. 1991; 66:15-17.